Real-time, deep-learning aided lensless microscope

JIMIN WU,[1] 🔲 **VIVEK BOOMINATHAN,[2]** 🔲 **ASHOK VEERARAGHAVAN,[2,3] AND JACOB T. ROBINSON[1,2,4,*]** 🔲

[1]*Department of Bioengineering, Rice University, Houston, Texas 77005, USA*
[2]*Department of Electrical and Computer Engineering, Rice University, Houston, Texas 77005, USA*
[3]*Department of Computer Science, Rice University, Houston, Texas 77005, USA*
[4]*Department of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA*
[*] jtrobinson@rice.edu

Abstract: Traditional miniaturized fluorescence microscopes are critical tools for modern biology. Invariably, they struggle to simultaneously image with a high spatial resolution and a large field of view (FOV). Lensless microscopes offer a solution to this limitation. However, real-time visualization of samples is not possible with lensless imaging, as image reconstruction can take minutes to complete. This poses a challenge for usability, as real-time visualization is a crucial feature that assists users in identifying and locating the imaging target. The issue is particularly pronounced in lensless microscopes that operate at close imaging distances. Imaging at close distances requires shift-varying deconvolution to account for the variation of the point spread function (PSF) across the FOV. Here, we present a lensless microscope that achieves real-time image reconstruction by eliminating the use of an iterative reconstruction algorithm. The neural network-based reconstruction method we show here, achieves more than 10000 times increase in reconstruction speed compared to iterative reconstruction. The increased reconstruction speed allows us to visualize the results of our lensless microscope at more than 25 frames per second (fps), while achieving better than 7 μm resolution over a FOV of 10 mm^2. This ability to reconstruct and visualize samples in real-time empowers a more user-friendly interaction with lensless microscopes. The users are able to use these microscopes much like they currently do with conventional microscopes.

1. Introduction

Fluorescence microscopes are critical in imaging the structures and functions of various biological samples. With advanced fabrication technologies, fluorescence microscopes have been scaled down, leading to new applications in the field of neuroscience and system biology [1–3]. However, traditional miniaturized microscopes are constrained by the fundamental trade-offs of lens-based imaging systems: the maximum achievable FOV, resolution, and light collection efficiency are dependent on the size of the lens.

Recent advancements computational imaging have enabled the development of lensless imaging systems that can potentially break free from these trade-offs [4–9]. Lensless microscopes, which replace lenses with light-modulating masks, offer substantial improvements in the tradeoffs between FOV, resolution and size [10–15]. Unlike traditional lens-based imaging systems that directly project a reproduction of a scene onto the sensor, lensless microscopes produce invertible transfer functions between the incident light field and the sensor measurements. These measurements on the sensor are highly multiplexed that require appropriate inverse algorithms to reconstruct focused images of the scene.

Despite the potential of lensless microscopes to overcome the trade-offs of lens-based microscopes, the adoption of lensless microscopes has been constrained by two interrelated factors. Firstly, fluorescence imaging on biological samples poses unique challenges for lensless

microscopes [16]. The emitted light from fluorophores is omnidirectional, and systems without lenses typically have low light-collection efficiency, which can compromise the reconstruction quality. Secondly, these computational microscopes cannot provide a real-time, live display option. As a result, crucial tasks in microscopy, such as selecting the sample's imaging field or setting the focus (or placing the sample at the right distance) become more of an art form than a repeatable engineering process.

The current lensless fluorescence microscopes are not ideal due to the mentioned limitations. They either provide near-real-time reconstructions at a long working distance with low light collection efficiency [13–15], or can operate at a close working distance with high light collection efficiency but require at least several minutes per frame for reconstruction [10,11]. Additionally, these limitations result in a lack of real-time feedback, which makes it difficult for individuals without prior knowledge of computational imaging to operate these microscopes. As a result, the adoption of these microscopes is significantly restricted.

In this work, we developed a compact and user-friendly lensless fluorescence microscope system that achieves real-time visualization with a wide FOV at a close working distance. This lensless microscope employs a contour-based phase mask designed to capture the texture frequencies commonly found in natural and biological samples [6,15], thus enabling high-quality reconstructions of dense and low-contrast biological samples. The real-time reconstruction visualisation was enabled by MultiFlatNet, a two-stage feed-forward deep neural network. With an easy assembly process, compact design, and better than 7 μm resolution across a 10 mm^2 FOV, our system provides an innovative solution for real-time fluorescence imaging.

Achieving high-quality imaging results in lensless microscopy largely depends on the reconstruction algorithms. Currently, numerous algorithms have been developed for image reconstruction in computational imaging systems. Classical methods for reconstruction are mostly iterative-based methods such as fast iterative shrinkage-thresholding algorithm (FISTA) [17] and the alternating direction multiplier method (ADMM) [18]. However, these methods are usually time-consuming and can be prone to noise amplification, which may compromise the accuracy of the reconstructed images. In order to improve the reconstruction quality, these methods incorporate pre-determined priors such as L1, Tikhonov, and total variation (TV). Inspite of these priors, existing reconstructions are noisy and display artifacts when reconstructing images of dense and low-contrast biological samples.

Deep-learning-based reconstruction methods have been applied to various computational imaging systems with improved reconstruction quality and speed [13,19–24]. This class of computational techniques that leverage machine learning provide a promising solution over traditional iterative-based methods. Nonetheless, the highly multiplexed data generated by the non-local optical encoders in the system pose a challenge to conventional data-driven methods such as convolutional neural networks, which have a limited receptive fields.

The MultiFlatNet we developed is a two-stage feed-forward network with a trainable inversion stage followed by a refining stage. The trainable inversion stage brings the highly multiplexed measurements back to image space, and is initialized using the captured PSFs. The refining stage leverages a fully convolutional network to address the spatial variance of the PSF and denoise the intermediate results. Typically, data-driven approaches require large-scale training data sets to generate accurate reconstructions on real captured data. To overcome this challenge, we developed an analytical forward model that utilizes a linear shift-variant model [10] to generate matched ground truth – lensless capture pairs. MultiFlatNet is then trained on this simulated dataset but generalizes well to real data as demonstrated in our experiments. The MultiFlatNet reconstruction achieves over 10,000 times faster reconstruction speed than existing reconstruction algorithms for spatially-varying PSFs based lensless imaging.

Our lensless microscope is a first-of-its-kind real-time, wide FOV lensless microscope. With the improved reconstruction speed, our lensless microscope can produce real-time visualizations

at a rate of more than 25 frames per second. This advancement makes the technology much more accessible and user-friendly, allowing individuals without any prior knowledge to use this microscope with ease. The users can simply place imaging targets under the lensless microscope, adjust the focal, and visualize real-time images, much like they would with conventional microscopes. Our user-friendly lensless microscope broadens access to scientists in need of compact fluorescence microscopes with a wide FOV and high resolution, but who are not familiar with computational imaging techniques.

2. Method

Lensless microscopes utilize light-modulating masks to capture a highly multiplexed version of the scene. In cases where the imaging distance is large, the mask-induced multiplexing can be considered consistent for the entire FOV, resulting a shift-invariant PSF. However, imaging at large distances sacrifices light collection efficiency, leading to reduced reconstruction quality and resolution. Moving the imaging sensor closer to the sample can significantly improve the light collection efficiency of the lensless microscopes, but it causes the PSF to change dramatically over the entire FOV, making the shift-invariant model unsuitable.

2.1. Shift-invariant forward model and reconstruction

In certain cases where the imaging distance is large, the PSF of the system can be regarded as being shift-invariant. The capture is considered as a global linear multiplexed version of the scene. The forward model can be described as the convolution of a single point-spread function (PSF) with the scene. This model is commonly used in existing lensless imaging systems [6,7,14,15]. In such cases, the forward model of the system can be described mathematically as follows:

$$\mathbf{b}(x, y) = p(x, y) * \mathbf{i}(x, y) + \mathbf{n}(x, y), \tag{1}$$

where $*$ represents convolution, $\mathbf{b}$ is the sensor measurement, p is the PSF, $\mathbf{i}$ is the scene, and $\mathbf{n}$ is the sensor and photon noise.

　　Lensless imaging systems commonly utilize optimization-based algorithms to solve the image reconstruction problem. Hand-picked regularizers are added to the optimization problem to minimize the noise amplification in the reconstruction. In traditional reconstruction algorithms, the scene $\mathbf{i}$ in Eq. (1) can be effectively solved by following regularized least squares problems:

$$\hat{\mathbf{i}} = \arg\min_{\mathbf{i} \geq \mathbf{0}} \|\mathbf{b} - p * \mathbf{i}\|_2^2 + \gamma \mathcal{R}(\mathbf{i}), \tag{2}$$

where $\mathcal{R}(\mathbf{i})$ is the regularization term, and the most commonly used regularization terms include L1, Tikhonov, and TV regularization terms. Assuming a shift-invariant PSF also allows for fast reconstruction using closed-form solutions. In particular, using Tikhonov regularization leads to a closed-form solution given by Wiener deconvolution and can be computed with the Fast Fourier Transform (FFT) as follows:

$$\hat{\mathbf{i}} = \mathscr{F}^{-1} \left(\frac{\mathscr{F}(p)^* \odot \mathscr{F}(\mathbf{b})}{|\mathscr{F}(p)|^2 + \gamma} \right), \tag{3}$$

where $\odot$ denotes Hadamard product, $(\cdot)^*$ is the complex conjugate operator, $\mathscr{F}$ denotes a Fourier transform and $\mathscr{F}^{-1}$ denotes an inverse Fourier transform, p is the captured single PSF when the point source is located at the center of the sensor, and γ is the regularization weight for Tikhonov regularization.

2.2. Shift-variant forward model and reconstruction

Lensless microscopes still face challenges in performing fluorescence imaging on biological samples. This is mainly due to the fact that the emitted light from fluorophores spreads in

all directions, and the light-collection efficiency of the lensless system is generally low. The light-collection efficiency of a lensless microscope has a inverse squared relationship with sample distance. Therefore, moving the imaging sensor closer to the sample can greatly improve the light collection efficiency of the lensless microscopes. However, when the object is in close proximity to the sensor, the point spread function changes significantly over the FOV, and the shift-invariant property of PSFs no longer holds. Figure 1(a) shows nine PSFs captured at various spatial locations of the point source, demonstrating the shift-varying property of the PSFs. When imaging at a closer distance, the low-frequency background resulting from autofluorescence and unfiltered excitation light can become more apparent. This low-frequency background does not fit in the physical model. In this case, the imaging system can be described mathematically [10,25]:

$$\mathbf{b}(x', y') = \mathbf{C} \sum_{x,y} p(x', y'; x, y) * \mathbf{i}(x, y) + \mathbf{n}(x', y') + \mathbf{g}(x', y'), \quad (4)$$

where $\mathbf{b}$ is the sensor measurement at the sensor position (x', y'), p is the PSF at the spatial location (x, y), $\mathbf{i}$ is the scene, $\mathbf{C}$ is the cropping function which accounts for the size of the sensor, $\mathbf{n}$ is the sensor and photon noise and $\mathbf{g}$ is the low-frequency background caused by filter autofluorescence or unfiltered excitation light.
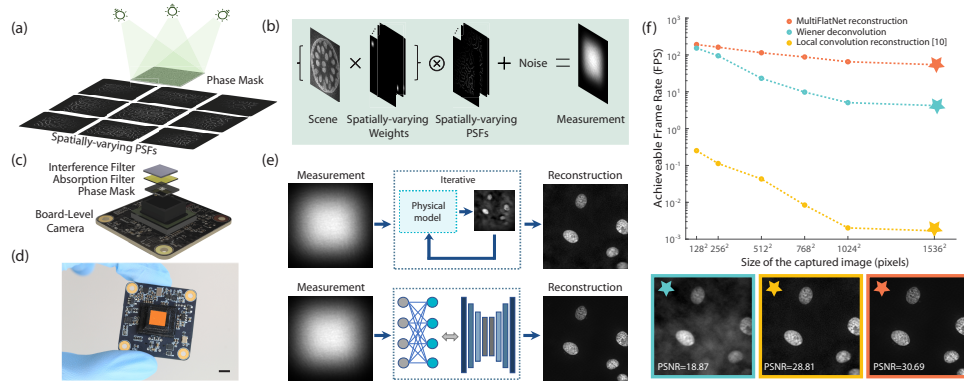

Fig. 1. Overview of the system. (a) Lateral scan of a point source through the FOV to capture the spatially-varying PSFs. (b) Imaging model of the system with shift-variant PSFs. (c) System diagram. The system contains a board-level camera, a phase mask, and a hybrid filter set. (d) A photo of an assembled prototype. Scale bar, 5 mm. (e) Comparison of traditional and proposed reconstruction approaches. (f) Achievable frame rate and reconstruction examples using various reconstruction methods. The stars indicate the actual size of the captured images in this study. Bottom shows example reconstructions using the three different reconstruction methods.

To calibrate at each resolvable location using the brute-force method, it would be necessary to sample the two-point resolution at the Nyquist rate. For a lensless microscope with wide FOV, this would involve obtaining over one million calibration samples on a single depth plane. Therefore, it is not feasible to measure the PSF for every spatial point due to the storage and computing source limitation. To address this issue, a "two-part model" was reported from previous studies [10,11]. This involves splitting the PSF measurements into two parts: a shift-invariant pattern generated by the phase mask, and a shift-varying term caused by the aberrations and sensor falloff. This allows us to perform sparse calibration and model-based interpolation to estimate the PSFs between calibration points. The estimated PSF from an arbitrary point can be described as:

$$p(x', y'; x, y) \approx \sum_{j} \alpha_j(x, y) \tilde{p}(x' + dx, y' + dy), \quad (5)$$

where $\tilde{p}(x', y')$ is the registered calibration measurements, $\alpha_j(x, y)$ is the weighting factor corresponding to the bilinear interpolation between the four nearest PSFs, and is determined by the distance to the *j*-th calibration measurement. The linear structure of this equation allows us to model the captured data without generating PSF for every spatial position. The imaging system can now be described as follows:

$$\mathbf{b}(x', y') = \mathbf{C} \sum_j [\alpha_j(-x', -y')\mathbf{i}(-x', -y')] * \tilde{p}(x', y') + \mathbf{n}(x', y') + \mathbf{g}(x', y'). \tag{6}$$

This local convolution model relies on precise measurement of the PSFs in space, and the lateral sampling distance is determined by Nyquist sampling theory [10]. In our system, we have set the spacing between each calibration measurement to be 0.5 mm, which has proven sufficient for accurate modeling. In our experiment, we calibrated a 13×13 grid (169 PSFs in total) on the imaging plane with 0.5 mm spacing (Supplement 1), covering a 6 mm × 6 mm area. Calibration images were captured using a 10 μm pinhole illuminated by a green LED (Thorlabs, M530L4) positioned behind an 80-degree holographic diffuser. Five measurements were taken for each calibration point and then averaged to reduce the noise.
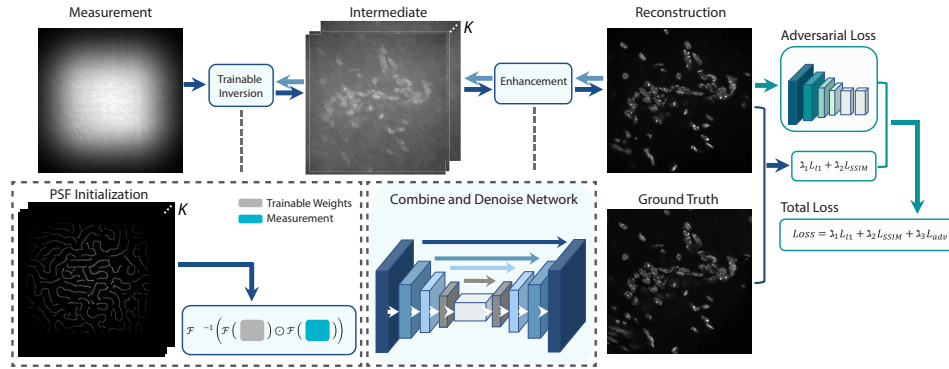

Fig. 2. The overall architecture of MultiFlatNet. The network contains two main parts: the sensor measurement is first transformed into the intermediate image space using a trainable inversion layer with five trainable weight matrices. After the inversion stage, a U-Net combines the intermediate images, enhances the quality, and produces the output image. A weighted combination of three losses is used as the loss function of our network, including an L1 loss, a structural similarity index measure (SSIM) loss, and an adversarial loss using a discriminator neural network.

Recovering the scene image from the measurement using shift-varying PSFs can still be posed as a convex optimization problem, with the local convolution model serving as the forward model. The scene **i** and background **g** can be jointly estimated by solving the following regularized least squares problem:

$$\hat{\mathbf{i}}, \hat{\mathbf{g}} = \arg\min_{\mathbf{i}, \mathbf{g} \geq \mathbf{0}} \|[\mathbf{C} \sum_j (\alpha_j \cdot \mathbf{i}) * \tilde{p} + \mathbf{g}] - \mathbf{b}\|_2^2 + \gamma \mathcal{R}(\mathbf{i}) \tag{7}$$

For the local convolution reconstruction used in this paper, we jointly solve for scene **i** and the background **g** by using FISTA. Without constraint to the background **g**, a trivial solution to Eq. (7) is **i** = 0 and **g** = **b**. We used 2D discrete consine transform (DCT) operator **D** to constrain the background estimation with **Dg** = 0 outside low frequency support. The constrain set to the DCT coefficients depends on the background level, and we set it to be zero outside the 5×5 lowest frequency components of **g** to prevent getting the unwanted trivial solution. In this paper, we specifically focused on reconstructions on a single depth plane. However, local convolution reconrtuation proposed in [10] can handle volumetric reconstructions.

2.3. Fast reconstruction using MultiFlatNet

Traditional reconstruction techniques often require computationally intensive and prohibitively slow iterative algorithms. Additionally, handpicked regularizers often rely on sparsity assumptions, which are not always suitable for imaging biological samples that tend to be of low contrast and not particularly sparse.

To overcome these challenges, we propose a deep learning-based approach called MultiFlatNet for real-time image reconstruction. Using a two-stage network (Fig. 2) structure similar to the structure proposed in [19,24], MultiFlatNet is specialized on lensless imaging systems with a larger support area in the captures, and incorporates a trainable adversarial loss which improves the reconstruction quality. The first stage is a trainable inversion process, which transforms the multiplexed sensor measurement to image-like intermediate reconstructions. The second stage is a neural network that combines the intermediate reconstructions to produce the final high-quality image.

Since the underlying image formation model is convolutional, the trainable inversion stage can be parameterized similarly to a deconvolution process as in Eq. (3). However, due to shift-variant PSFs, a single deconvolution process will be insufficient, and having separate ones for every 13×13 (total of 169) PSFs will be computationally intensive and slow. Instead, we choose $K << 169$ deconvolution layers to map the sensor measurement to K image-like intermediate reconstructions. The K intermediates are concatenated along the channel dimension at the end of the first stage.

The parameterization of the trainable inversion stage is in the form of Hadamard product in the Fourier domain given as:

$$\hat{I}_k = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{b}) \odot \mathcal{F}(W_k)), k = 1, 2, \ldots, K, \tag{8}$$

where $\hat{I}_k$ is the output of each channel in the intermediate stage and W_k is the learnable filter for that channel, $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the FFT and the inverse FFT operations, and $\odot$ is the Hadamard product. We show that $K = 5$ is sufficient to achieve high-quality reconstruction over 10 mm^2 FOV at real-time speeds (Supplement 1). The trainable weights W_k are initialized using K evenly distributed PSFs as $\mathcal{F}^{-1}\left(\frac{H_k^*}{|H_k|^2+\gamma}\right)$, where H_k is the Fourier transform of the corresponding PSF and γ is a regularization term. Supplement 1 shows the $K = 5$ evenly distributed PSFs chosen based on the achievable FOV among the 13×13 calibration positions.

In the second stage of MultiFlatNet, we use a fully convolutional network to map the K-channel intermediate outputs to the final high-quality image reconstruction. We choose U-Net-based [26] architecture as our network due to its multi-resolution structure that has proven very successful in image-to-image translation problems. The kernel size was fixed at 3×3, and the number of filters is gradually increased from 32 to 256 in the encoder and reduced back to 32 in the decoder. The final layer produces a single channel image, which is our final output image.

Having appropriate loss functions is important for the system to produce high-quality reconstructions. In the MultiFlatNet reconstruction, an adversarial loss was added to ensure that the distribution of the reconstructed output closely matches that of the ground truth images. We used a discriminator neural network [19,23,27] with 4 layers of 2-strided convolution followed by batch normalization and ReLU activation function for the adversarial loss. The total loss we used was a weighted combination of L1 loss, structural similarity index measure (SSIM) loss and adversarial loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{L1}} + \lambda_2 \mathcal{L}_{\text{SSIM}} + \lambda_3 \mathcal{L}_{\text{adv}}, \tag{9}$$

where λ_1, λ_2 and λ_3 are weights assigned to each term.

2.4.   System hardware development

Our lensless microscope is a phase-mask based lensless imaging system and consists of three main components: a monochromatic imaging sensor (Imaging Source, SMM 37UX178-ML, 2.4 μm pixels), a contour-based phase mask [15], and a hybrid filter set (Fig. 1(c)). The contour PSF is generated using Perlin noise with Canny edge detection applied. To design the phase mask, we used a phase retrieval algorithm [6] with a feature size of 6 μm. The phase mask was fabricated using a two-photon lithography system (Nanoscribe, Photonic Professional GT) on a 700 μm thick fused silica substrate. A photoresist (Nanoscribe, IP-Dip) was used for fabrication. The size of the fabricated phase mask is 3 mm × 3 mm (Supplement 1) to ensure a high light throughput for low-contrast fluorescent samples. The designed phase mask has a 1 mm focal length, with a 2 mm mask-to-sensor distance and a 2 mm mask-to-scene distance. The designed distances are much closer than our previous design in [15] and achieved much higher light collection efficiency. The design and fabrication details of the phase mask can be found in the Supplement 1. To achieve sufficient imaging contrast for biological samples, we used a hybrid filter set consisting of an adsorptive filter (Kodak, Weatten 12) placed under an interference filter (Chroma, ET 525/50m), similar to the design in [15,28–30]. This hybrid filter set efficiently removes the excitation light. A 3D-printed housing (MJP 2500) is used to hold the phase mask and hybrid filter set on top of the imaging sensor. Compare to our previous design [15], this new lensless microscope features a larger phase mask with smaller fabrication feature size. This enables it to capture high-quality images at closer distances with reconstruction algorithms using spatially-varying PSFs. This results in improved light throughput and image quality.

2.5.   Training data generation and implementation details

Collecting a large-scale dataset for lensless imaging systems with matched ground truth is challenging, and supervised training generally requires a large-scale labeled dataset. To address this challenge, we developed a forward model simulator using 169 calibrated PSFs on a single imaging plane based on the local convolution forward model introduced in Section 2.2. The simulator was used to generate measurements from the open-source microscopy data [31–33], widefield microscopy data captured at Rice University, and widefield microscopy data captured at University of Texas at Austin. Specifically, we selected 7000 images from the open-source dataset and 3000 captured widefield microscopy data for the training process. We randomly selected 9000 images for training and reserved 1000 images for testing. All ground truth images were resized to 768×768 pixels. The simulated captures were generated by the forward model simulator using 169 PSFs calibrated at approximately 1.5 mm imaging distance. We then added a realistic level of Gaussian noise to the simulated captures. The parameters for the Gaussian noise for each simulated capture are randomly selected within the range estimated by experimental measurements. The measurement dimension of the system is 2048×3072, and we cropped them to 1536×1536 during the training process. A small batch size was used due to memory constraints. We trained for 100 epochs using a learning rate of 10^{-4} with the Adam optimizer [34].

3.   Experiments and results

3.1.   System characterization

We characterized the spatial resolution of the lensless microscope using a negative 1951 USAF resolution target (Edmund Optics 59-204) with an added fluorescent background. Specifically, our system can achieve a resolution of < 10 μm at imaging depths up to 3.5 mm. For each imaging depth, 10 images were captured at 20 ms each, and averaged to remove noise. The excitation light was provided by an external near-collimated 470 nm LED (Thorlabs, M470L3) with an incorporated excitation filter (Thorlabs, MF 469-35). Images captured at different depths

were characterized in Fig. 3(f), which demonstrates the ability of our system to maintain high resolution across a wide imaging depth range. Figure 3(e) shows an example reconstructed USAF target at a depth of 1.5 mm, and could clearly resolve the features in group 6 element 2, which indicates a ~7 μm resolution. We also prepared a sample by spreading 15 μm fluorescent beads on a glass slide and captured the image at 1.5 mm depth. The full width at half maximum (FWHM) of the lateral spread of the beads was around 15 μm (Fig. 3(c)), indicating that our system can accurately reconstruct small samples. Visualization 1 shows a screen recording of the real-time captures and reconstructions of this slide moving in the FOV.
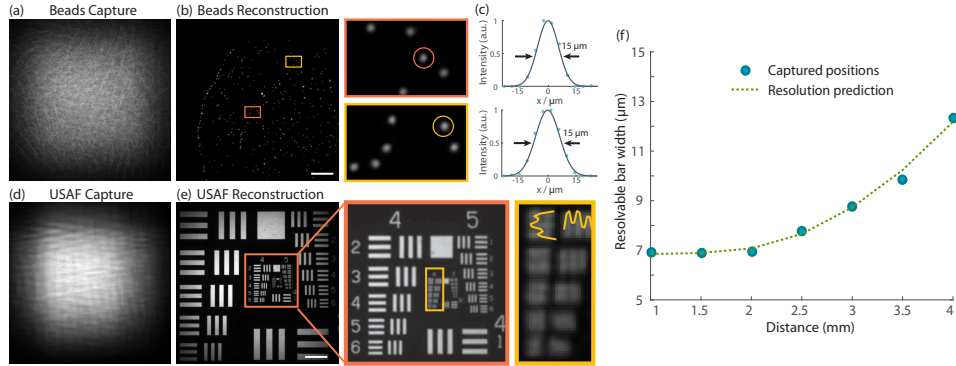
Fig. 3. System Characterization. (a) Experimental capture of 15 μm green fluorescent beads spread on a glass slide. (b) The reconstruction of 15 μm green fluorescent beads spread on a glass slide. Single beads are clearly visible in the small region zoom-ins. Scale bar, 500 μm. (c) The lateral profile of two representatives reconstructed beads in the circle in panel b. (d) Experimental capture of a negative USAF target with an added fluorescent background. (e) The reconstruction of the negative USAF target. Zoom-in shows groups 4,5 and 6. Group 6 element 2 (< 7 μm features) can be resolved in the reconstructed image. Scale bar, 500 μm. (f) Resolution evaluation at different imaging depths.

3.2. Reconstruction performance on simulated measurements

We first evaluated the performance of the MultiFlatNet on simulated captures. Reconstructed images were cropped to a size of 768×768 pixels. The results shown that the MultiFlatNet achieved a high reconstruction quality while being more than 10000 times faster than the traditional iterative-based method. We evaluated the performance of our trained model on a held-out testing dataset of 1000 images, and selected representative results are shown in Fig. 4. We compared our MultiFlatNet reconstruction results (Fig. 4(d)) with those obtained using Wiener deconvolution using a single PSF (Fig. 4(b)) and FISTA local convolutional reconstruction using 169 PSFs (Fig. 4(c)), as described in Section 2.2. Although FISTA local convolution reconstruction results exhibit a comparable level of detail as the MultiFlatNet reconstruction results, there are noticeable reconstruction artifacts (Fig. 4 zoom-ins), especially in densely labeled samples. Table 1 presents the average metrics of different reconstruction methods, including peak signal-to-noise ratio (PSNR), runtime, and achievable frame rate. The MultiFlatNet reconstruction outperforms traditional reconstruction methods, providing fast and accurate reconstructions that enable high frame rate real-time visualization. The runtime of Wiener deconvolution was tested on an 8-core processor (AMD Ryzen 7 3700X, 3.59GHz) with 64 GB RAM, and the other two methods were tested on Nvidia GeForce RTX 2070 GPU.
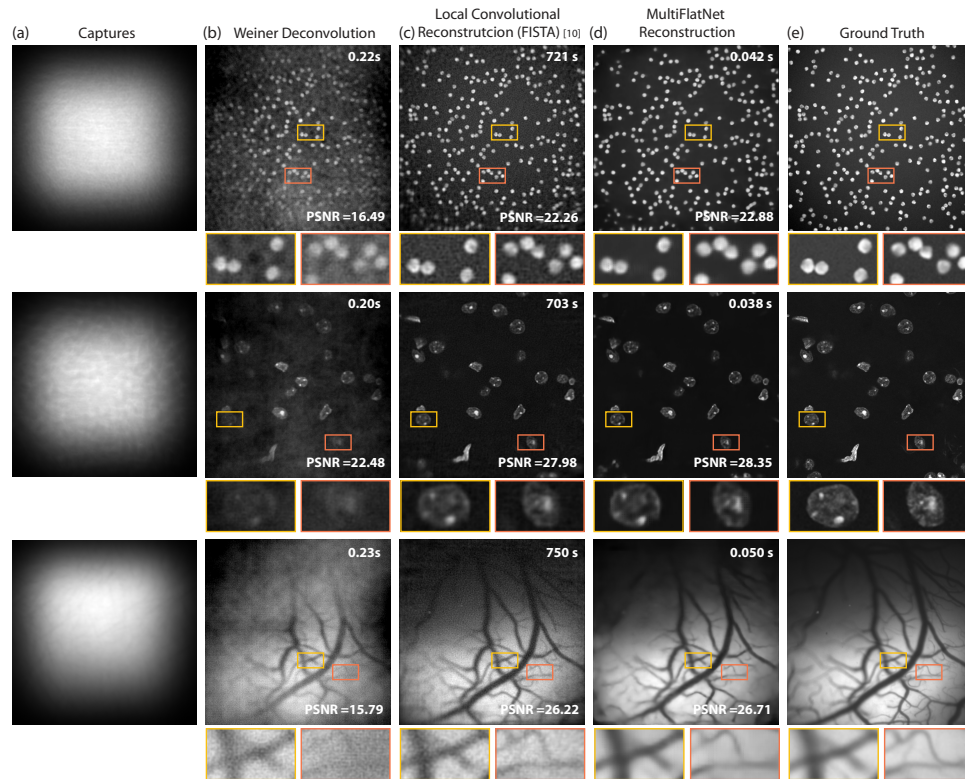
Fig. 4. Reconstruction results of simulated captures using various approaches. (a) Simulated captures using forward model simulation. (b) Wiener deconvolution reconstructions using a single PSF captured when the point source was located at the center of the FOV (c) Local convolutional reconstructions (FISTA) using 169 PSFs. (d) Reconstruction using the MultiFlatNet. The network reconstruction achieves better and faster results than other traditional methods. (e) Ground truth images.

Table 1. Average metrics on the simulated testing dataset (capture size 1536 × 1536 pixels).

Method	PSNR (in dB)	Runtime (s)	Frame Rate (fps)
Wiener Deconvolution	17.73	0.214	4.7
Local Convolution Reconstruction [10]	25.45	725	0.0014
MultiFlatNet	26.09	0.039	25

3.3. Reconstruction performance on real measurements

Despite being trained on simulated data, the MultiFlatNet reconstruction is capable of generating high-quality reconstructions of biological samples captured by our lensless microscope. To demonstrate the capability of our system in imaging biological samples, we imaged several different samples using our system. We reconstructed the captures using Wiener deconvolution (Fig. 5(b)), FISTA local convolutional reconstruction (Fig. 5(c)), and the MultiFlatNet reconstruction (Fig. 5(d)). We compared these results with the ground truth images (Fig. 5(e)), which were obtained using a widefield microscope with a 4× microscope objective (Nikon Fluor). Our first sample was a slide of Convallaria majalis (lily of the valley) stained with Acridine-Orange. We captured five images at 50 ms each, and averaged them to reduce noise (Fig. 5(a)). The zoom-ins of the convallaria reveal that our system can accurately reconstruct the large circular plant cells with sizes around 10 μm. Visualization 2 shows a screen recording of the real-time captures and reconstructions of the Convallaria majalis slide moving in the FOV.
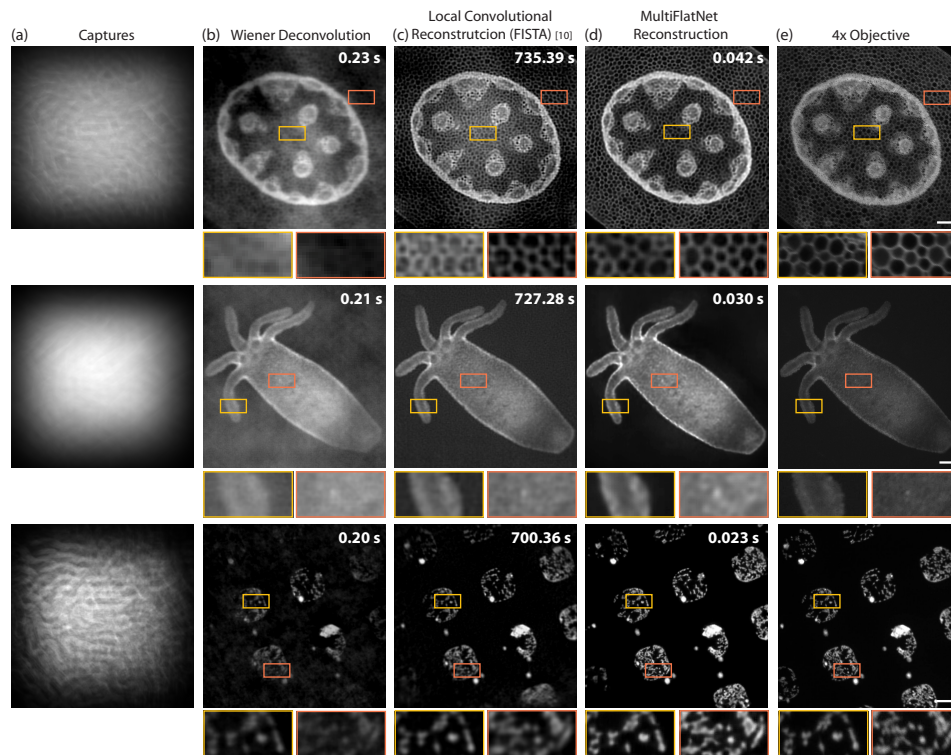

Fig. 5. Reconstruction results of real captures using various approaches. (a) Real captures of our system. (b) Wiener deconvolution reconstructions using a single PSF captured when the point source was located at the center of the FOV (c) Local convolutional reconstructions (FISTA) using 169 PSFs. (d) Reconstruction using the MultiFlatNet. The network reconstruction achieves better and faster results than other traditional methods. (e) Ground truth images. Scale bar, 200 μm.

To further demonstrate our system's capability to accurately reconstruct biological samples, we conducted experiments on a Watermelon Hydra Vulgaris expressing GFP in the ectoderm and RFP in the endoderm. The Hydra was treated with 4% formaldehyde diluted in Hydra media from a 16% formaldehyde solution (Thermo Fisher 28906) for 30 s to ensure that they stayed in the same position over time. Then the Hydra was immediately inserted into a 100 μm-thick

microfluidic chamber [35] for imaging. The samples were captured by our system immediately after being prepared, and the ground truth images were obtained within 30 minutes of capturing the samples. In this experiment, we used a blue LED (Thorlabs, M470L3) as the excitation light source and focused solely on imaging the GFP in the ectoderm. To remove random noise, we took the average of five captured images at 200 ms each. In the reconstructed images, we were able to identify some cellular-level structures of around 10 μm size using both the local convolutional reconstruction and MultiFlatNet reconstruction methods with comparable quality, while the MultiFlatNet reconstruction was significantly faster.

To show the system's capability to resolve individual cells in a dense biological sample, we imaged a sample containing spiking human embryonic kidney (HEK) cells with 3D printed patterns. Spiking HEK 293 cells [36] are cultured in DMEM-F12 (Lonza) supplemented with 10% FBS (Gibso), and 1% penicillin/streptomycin (Lonza). The coverslip holding the cells is made on a photolithographically patterned 12 mm glass coverslip with polydimethylsiloxane for 300 μm circles every 300 μm [37,38]. The cells are replated onto the coverslip (around 10,000 cells per coverslip) and grow for 2 days, resulting in isolated HEK 293 cells colonies of defined size and geometry. The cells are incubated with 2 μM Calcein-AM 30 minutes before recordings. The coverslip is transferred into a petri dish containing 1 mL of PBS for imaging. The samples were captured by our system immediately after being prepared, and the ground truth images were taken immediately after our system (within 30 minutes). The captured image is the average of five captures at 100 ms each to remove noise. Our system successfully identified the single cells in the colonies with similar quality to the ground truth images, with the size of the HEK cells around 13 μm.

These results demonstrate that our system achieves cellular-level resolution on various complex biological samples, and the reconstruction quality is comparable to that of a conventional widefield microscope with a 4× objective. The total frame rate is determined by the sum of the exposure time and reconstruction time. In our studies the exposure times for imaging biological samples are dependent on the fluorescence level of the sample, which is comparable to the exposure times we observed with 4× objective lenses. Moreover, the MultiFlatNet allows for fast and accurate reconstruction on real samples, with comparable quality to the reconstructions obtained from simulated data. This observation confirms that our forward model simulation accurately produces measurements that closely resemble those obtained in real-world measurements.

3.4. User study

Taking advantage of the fast reconstruction algorithm, we developed a user interface (Supplement 1) for our lensless microscope, which displays real-time captures and reconstructions. We conducted a usability study with volunteers who were unfamiliar with computational microscopes. In this experiment, the lensless microscope was mounted on a motorized linear translational stage (Thorlabs, LNR502E), and the sample slide was placed on a 3D-printed sample holder (Formlabs, Form 3) with a light source (Thorlabs, M470L3) placed beneath it (Supplement 1). A total of eight users participated in this study, all of whom were familiar with conventional microscopes, but not with lensless microscopes. After receiving a brief introduction on how to use the lensless microscope, users were asked to image two fluorescence targets using a conventional microscope and the lensless microscope with two different reconstruction methods. We recorded and compared the time taken to complete each imaging task for all users. When using the lensless microscope, the users performed imaging tests using both the MultiFlatNet reconstruction with real-time feedback, and the Wiener deconvolution without real-time display. Here we did not compare to FISTA local convolution reconstruction because it is extremely time-consuming and difficult for non-experienced users to use.

Figure 6 shows the testing process and outcomes of the usability study. The timer began once the users placed the fluorescence target on the sample holder and ended when the target
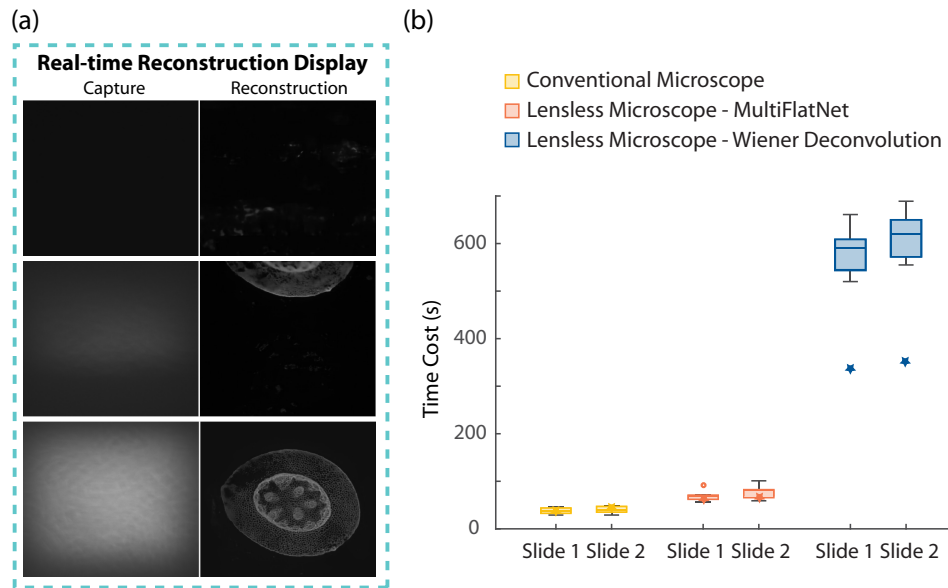
(a)

(b)

Fig. 6. User study. (a) Examples of real-time capture and reconstruction displays for targets located outside the FOV, at the FOV's edge, and at its center. (b) The time costs when imaging two fluorescent samples using a conventional microscope and our lensless microscope. The time costs for the lead author of this paper is indicated by star. Other subjects in the study were not involved in the research.

was aligned to the center of the FOV and focused optimally. Our lensless microscope achieves a comparable level of usability to a conventional microscope by utilizing the MultiFlatNet reconstruction with real-time feedback. Conversely, when using Wiener deconvolution without real-time feedback, it is challenging for non-experienced users to identify the correct location and depth of the target. The time required for users to image the sample using Wiener deconvolution is significantly higher (600 s) compared to the MultiFlatNet reconstruction (72 s) and the conventional microscope (39 s). In addition, we evaluated the time costs for an experienced user (lead author of this paper), and denoted them with stars in Fig. 6(b). Our findings indicate that for the experienced user, the time cost is significantly lower when using the lensless microscope with Wiener deconvolution than for non-experienced users. However, there was no significant difference in time cost when using either the conventional microscope or the lensless microscope with the MultiFlatNet reconstruction. The time required for imaging with our lensless microscope using the MultiFlatNet reconstruction is only slightly higher compared to using a conventional microscope. According to user feedback, this time difference was primarily due to the motorized translation stage we used. Most of users were not familiar with the software-controlled translation stage, and it generally moves slower than the manual stages in conventional microscopes. As a result, it was harder for users to find the position of the target using motorized translation stages than manual stages. In addition to the motorized translation stage, the close working distance of the lensless microscope could also be a contributing factor to the longer time required. The close working distance makes it more challenging for users to precisely position the target directly under the imaging sensor area. Despite this challenge, the usability study indicates that users without prior knowledge and experience can easily use this lensless microscope, and the operation process is similar to that of current conventional microscopes.

4. Conclusion and discussion

In conclusion, we demonstrated a compact, user-friendly lensless fluorescence microscope that achieves real-time reconstruction and visualization using a trainable deep neural network. To the best of our knowledge, this lensless microscope is the first real-time lensless microscope with usability comparable to conventional lens-based microscopes, with the benefit of small size and superior FOV. While the reconstruction network is currently trained on a single depth plane, there is potential to extend the capability to multiple planes and even perform 3D reconstructions. Once the network is trained, it is able to provide a significant increase in reconstruction speed, exceeding 10,000 times that of an iterative-based reconstruction method using spatially-varying PSFs. Due to the fact that our trainable inversion stage is initialized using only five PSFs, there may be some degradation in reconstruction quality at the edges of the FOV. Nevertheless, the quality of our reconstruction is still comparable to that achieved with traditional reconstruction methods using spatially-varying PSFs. While this could be improved by using more densely calibrated PSFs in more channels for the trainable inversion stage, such an approach would increase runtime and memory requirements, preventing the device from high frame rate real-time usage. The Supplement 1 provides a comparison of run time and image quality using different numbers of PSFs for this reconstruction method. It should be noted that, like the most data-driven method, the quality of the network-based reconstruction depends on the training dataset used. To ensure good quality reconstructions, the training dataset must be diverse enough. Transfer learning can be employed on a specific dataset to improve the reconstruction quality for applications with known target distribution. This low-cost, compact, and small-size microscope with real-time visualization unlocks new possibilities for research and medical applications in laboratory settings and resource-limited areas. It can be used to study neural circuits in living organisms and for quick diagnosis and monitoring of treatment efficacy in various medical conditions. Its portability and affordability make it accessible to a broader range of researchers and healthcare professionals, potentially positively impacting multiple fields.

Disclosures. The authors declare no conflicts of interest.

Data availability. Designed user interface, reconstruction algorithm with pretrained network, and sample data are available at [39].

Supplemental document. See Supplement 1 for supporting content.

References

1. O. Skocek, T. Nöbauer, L. Weilguny, F. Martínez Traub, C. N. Xia, M. I. Molodtsov, A. Grama, M. Yamagata, D. Aharoni, D. D. Cox, P. Golshani, and A. Vaziri, "High-speed volumetric imaging of neuronal activity in freely moving rodents," Nat. Methods **15**(6), 429–432 (2018).
2. K. K. Ghosh, L. D. Burns, E. D. Cocker, A. Nimmerjahn, Y. Ziv, A. E. Gamal, and M. J. Schnitzer, "Miniaturized integration of a fluorescence microscope," Nat. Methods **8**(10), 871–878 (2011).
3. M. L. Rynes, D. A. Surinach, S. Linn, M. Laroque, V. Rajendran, J. Dominguez, O. Hadjistamoulou, Z. S. Navabi, L. Ghanbari, G. W. Johnson, M. Nazari, M. H. Mohajerani, and S. B. Kodandaramaiah, "Miniaturized head-mounted microscope for whole-cortex mesoscale imaging in freely behaving mice," Nat. Methods **18**(4), 417–425 (2021).
4. V. Boominathan, J. T. Robinson, L. Waller, and A. Veeraraghavan, "Recent advances in lensless imaging," Optica **9**(1), 1 (2022).
5. M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "FlatCam: thin, lensless cameras using coded aperture and computation," IEEE Trans. Comput. Imaging **3**(3), 384–397 (2017).

6. V. Boominathan, J. K. Adams, J. T. Robinson, and A. Veeraraghavan, "PhlatCam: designed phase-mask based thin lensless camera," IEEE Trans. Pattern Anal. Mach. Intell. **42**(7), 1618–1629 (2020).
7. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," Optica **5**(1), 1 (2018).
8. W. Chi and N. George, "Optical imaging with phase-coded aperture," Opt. Express **19**(5), 4294 (2011).
9. W. Chi and N. George, "Phase-coded aperture for optical imaging," Opt. Commun. **282**(11), 2110–2117 (2009).
10. G. Kuo, F. Linda Liu, I. Grossrubatscher, R. Ng, and L. Waller, "On-chip fluorescence microscopy with a random microlens diffuser," Opt. Express **28**(6), 8384 (2020).
11. J. K. Adams, V. Boominathan, B. W. Avants, D. G. Vercosa, F. Ye, R. G. Baraniuk, J. T. Robinson, and A. Veeraraghavan, "Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope," Sci. Adv. **3**(12), e1701548 (2017).
12. F. Tian, J. Hu, and W. Yang, "GEOMScope: large field-of-view 3D lensless microscopy with low computational complexity," Laser & Photonics Reviews **15**, 1863–1880 (2021).
13. Y. Xue, Q. Yang, G. Hu, K. Guo, and L. Tian, "Deep-learning-augmented computational miniature mesoscope," Optica **9**(9), 1009 (2022).
14. Y. Xue, I. G. Davison, D. A. Boas, and L. Tian, "Single-shot 3D wide-field fluorescence imaging with a computational miniature mesoscope," Sci. Adv. **6**(43), eabb7508 (2020).
15. J. K. Adams, D. Yan, J. Wu, V. Boominathan, S. Gao, A. V. Rodriguez, S. Kim, J. Carns, R. Richards-Kortum, C. Kemere, A. Veeraraghavan, and J. T. Robinson, "In vivo lensless microscopy via a phase mask generating diffraction patterns with high-contrast contours," Nat. Biomed. Eng **6**(5), 617–628 (2022).
16. A. Greenbaum, W. Luo, T.-W. Su, Z. Göröcs, L. Xue, S. O. Isikman, A. F. Coskun, O. Mudanyali, and A. Ozcan, "Imaging without lenses: achievements and remaining challenges of wide-field on-chip microscopy," Nat. Methods **9**(9), 889–895 (2012).
17. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. **2**(1), 183–202 (2009).
18. S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," FNT Mach. Learn. **3**(1), 1–122 (2010).
19. S. S. Khan, V. Sundar, V. Boominathan, A. Veeraraghavan, and K. Mitra, "FlatNet: towards photorealistic scene reconstruction from lensless measurements," ArXiv, ArXiv:2010.15440 (2020).
20. X. Pan, X. Chen, S. Takeyama, and M. Yamaguchi, "Image reconstruction with transformer for mask-based lensless imaging," Opt. Lett. **47**(7), 1843 (2022).
21. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," Opt. Express **27**(20), 28075 (2019).
22. J. Yang, X. Yin, M. Zhang, H. Yue, X. Cui, and H. Yue, "Learning image formation and regularization in unrolling AMP for lensless image reconstruction," IEEE Trans. Comput. Imaging **8**, 479–489 (2022).
23. S. S. Khan, A. V. R. , V. Boominathan, J. Tan, A. Veeraraghavan, and K. Mitra, "Towards photorealistic reconstruction of highly multiplexed lensless images," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, (IEEE, 2019), pp. 7859–7868.
24. K. Yanny, K. Monakhova, R. W. Shuai, and L. Waller, "Deep learning for fast spatially varying deconvolution," Optica **9**(1), 96 (2022).
25. K. Yanny, N. Antipa, W. Liberti, S. Dehaeck, K. Monakhova, F. L. Liu, K. Shen, R. Ng, and L. Waller, "Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy," Light: Sci. Appl. **9**(1), 171 (2020).
26. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," ArXiv,ArXiv:1505.04597 (2015).
27. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, vol. 27 Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds. (Curran Associates, Inc., 2014).
28. A. Ozcan and U. Demirci, "Ultra wide-field lens-free monitoring of cells on-chip," Lab Chip **8**(1), 98–106 (2008).
29. C. Richard, A. Renaudin, V. Aimez, and P. G. Charette, "An integrated hybrid interference and absorption filter for fluorescence detection in lab-on-a-chip devices," Lab Chip **9**(10), 1371 (2009).
30. M. B. Bouchard, V. Voleti, C. S. Mendes, C. Lacefield, W. B. Grueber, R. S. Mann, R. M. Bruno, and E. M. C. Hillman, "Swept confocally-aligned planar excitation (SCAPE) microscopy for high-speed volumetric imaging of behaving organisms," Nat. Photonics **9**(2), 113–119 (2015).
31. S. Arslan, T. Ersahin, R. Cetin-Atalay, and C. Gunduz-Demir, "Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images," IEEE Trans. Med. Imaging **32**(6), 1121–1131 (2013).
32. Y. Zhang, Y. Zhu, E. Nichols, Q. Wang, S. Zhang, C. Smith, and S. Howard, "A Poisson-Gaussian denoising dataset with real fluorescence microscopy images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (IEEE, 2019), pp. 11702–11710.
33. V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation," Nat. Methods **9**(7), 637 (2012).
34. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," ArXiv, ArXiv:1412.6980 (2017).
35. K. N. Badhiwala, D. L. Gonzales, D. G. Vercosa, B. W. Avants, and J. T. Robinson, "Microfluidics for electrophysiology, imaging, and behavioral analysis of *Hydra*," Lab Chip **18**(17), 2523–2539 (2018).

36. J. Park, C. A. Werley, V. Venkatachalam, J. M. Kralj, S. D. Dib-Hajj, S. G. Waxman, and A. E. Cohen, "Screening Fluorescent Voltage Indicators with Spontaneously Spiking HEK Cells," PLoS One **8**(12), e85221 (2013).
37. Y. Zhu, D. Sazer, J. S. Miller, and A. Warmflash, "Rapid fabrication of hydrogel micropatterns by projection stereolithography for studying self-organized developmental patterning," PLoS One **16**(6), e0245634 (2021).
38. B. Grigoryan, D. W. Sazer, A. Avila, J. L. Albritton, A. Padhye, A. H. Ta, P. T. Greenfield, D. L. Gibbons, and J. S. Miller, "Development, characterization, and applications of multi-material stereolithography bioprinting," Sci. Rep. **11**(1), 3171 (2021).
39. J. Wu, V. Boominathan, A. Veeraraghavan, and J. T. Robinson, "Real-time lensless microscope," Github, 2023, https://github.com/JiminWu/Real-time-lensless-microscope.